



رسم درخت فیلوژنتیکی بر مبنای بیسین (Bayesian)

آزاده حبیبی^۱، بنفشه صفایی فراهانی^۲، رضا مستوفی زاده قلمفرسا^۳

^۱ استادیار، گروه تنوع زیستی، پژوهشگاه علوم و تکنولوژی پیشرفته و علوم محیطی، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان،

a.habibi@kgut.ac.ir

^۲ استادیار، بخش تحقیقات گیاهپزشکی، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان فارس، سازمان تحقیقات، آموزش و ترویج کشاورزی، شیراز،

bsafaie@shirazu.ac.ir

^۳ استاد، بخش گیاه پزشکی، دانشگاه شیراز، rmostofi@shirazu.ac.ir

چکیده

هدف اصلی تحقیقات فیلوژنتیکی بازسازی تاریخچه تکاملی موجودات است. ارتباطات خویشاوندی بین موجودات مختلف از طریق رسم دیاگرام‌های درختی به نام درخت‌های فیلوژنی توصیف می‌شود. از بین روش‌های متداول روش بیسین روشی قدرتمند برای رسم درختان فیلوژنی است. در این روش با توجه به داده‌های مشاهده شده، احتمال صحت درخت که به آن احتمال پس‌آیند گفته می‌شود، بررسی می‌گردد. این روش جوهره‌ای از روش بیشینه درست‌نمایی است که به جای جستجوی درختی که احتمال دیدن داده‌های موجود را حداکثر می‌کند، با توجه به داده‌های موجود به دنبال درختی با بیشترین احتمال می‌گردد. از مزایای این روش نسبت به روش بیشینه درست‌نمایی این است که الگوریتم استفاده شده برای هر یافته پشته‌ای آماری ارائه می‌کند که این پشته‌ای احتمال واقعی است در حالی که در روش بیشینه درست‌نمایی مقادیر بوت استرپ، احتمال نیستند.

واژه‌های کلیدی: احتمال پس‌آیند، بیسین، روش‌های فیلوژنتیکی

مقدمه

فیلوژنی تاریخچه تکاملی گروهی از نهاده‌ها است. هدف از فیلوژنی توصیف ارتباطات تکاملی از لحاظ خویشاوندی نسبی با جد مشترک است. این ارتباطات به شکل دیاگرامی شاخه شاخه یا درختی نشان داده می‌شود که شاخه‌های آن از طریق گره‌ها به هم وصل هستند [1]. در روش‌های سنتی واکاوی‌های فیلوژنتیکی از روی مشخصات ریخت‌شناختی صورت می‌گرفت ولی امروزه با توجه به حجم بالای داده‌های مولکولی موجود، نقطه آغاز واکاوی عمده‌تر دسترسی به اطلاعات توالی (توالی‌های اسید نوکلئیک یا پروتئین) است. با این وجود، دسترسی به چنین منبع قابل اعتمادی از شواهد به معنی پایان مشکلات و مسائل تحقیق فیلوژنتیکی نیست. بلکه جنبه‌های محاسباتی اهمیت می‌یابند. روش‌های اصلی رسم درخت‌های فیلوژنی

شامل روش‌های مبتنی بر فاصله^۱ و روش‌های مبتنی بر صفت^۲ است. روش‌های مبتنی بر فاصله مانند روش الحاق مجاور^۳، فاصله‌های دوتایی بین توالی‌ها را محاسبه کرده و توالی‌های دارای بیشترین تشابه را در یک گروه قرار می‌دهد. به عبارت دیگر فاصله بین هر جفت توالی محاسبه می‌شود و از ماتریس فاصله به دست آمده برای ساخت درخت استفاده می‌شود. این روش ساده و سریع است و برای بررسی‌های اولیه مفید است اما برای استفاده از درخت نهایی باید آن را با نتایج سایر روش‌ها مقایسه کرد. در مقابل، روش‌های مبتنی بر صفات مانند روش‌های پارسمونی^۴، بیشینه درست‌نمایی^۵ و بیسین^۶ هستند. روش‌های مبتنی بر صفات، به طور هم زمان کلیه توالی‌های مورد بررسی را مقایسه می‌کند در حالی که برای ساخت درخت هر بار یک صفت را در نظر گرفته و یک امتیاز به هر درخت می‌دهد. در روش بیشینه پارسمونی امتیاز درخت کمترین تعداد تغییرات است. در روش بیسین این امتیاز احتمال پس‌آیند^۷ است. از لحاظ تئوری درختی که بهترین امتیاز را دارد از طریق مقایسه کلیه درختان ممکن به دست می‌آید. در عمل به دلیل تعداد بالای درختان ممکن این کار غیر ممکن است و برای حل این مشکل از الگوریتم‌های جستجوی هورستیک استفاده می‌شوند. این روش‌ها درختی اولیه با یک الگوریتم سریع تولید کرده و سپس برای بهبود امتیاز درخت تنظیمات مجدد انجام می‌دهند. روش‌های بیشینه درست‌نمایی و بیسین بر مبنای مدل هستند اما روش‌های پارسمونی بر مبنای مدل نیستند [2]. روش‌های آماری برای بازسازی فیلوژنی سعی بر محاسبه احتمال درخت‌های فیلوژنتیکی مختلف دارند و به این منظور باید از یک مدل تکامل توالی استفاده کنند. مدل تکامل توالی این واقعیت را در نظر می‌گیرد که احتمال وقوع تمام جهش‌ها در یک توالی یکسان نیست [1].

¹ distance based methods

² Character based methods

³ neighbor joining (NJ)

⁴ parsimony

⁵ maximum likelihood

⁶ Bayesian

⁷ posterior probability